Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 46 No. 11
November 2025

Comparative Analysis of Choquet Integral and Takagi–Sugeno for Ensemble Fusion in Deepfake Detection

P Noelien Selvarani[1], M Selvarathi[2*]

[1] Research Scholar, School of Sciences, Arts and Media, Karunya Institute of Technology and Sciences, Coimbatore, India. Email: noelienselvarani@gmail.com

[2] Assistant Professor, Department of Mathematics, Division of Food Processing Technology, Karunya Institute of Technology and Sciences, Coimbatore, India. Email: selvarathi.maths@gmail.com

Abstract

Deepfakes pose a significant threat to digital trust and online media authenticity. Although deep learning-based classifiers such as Xception and EfficientNet achieve strong detection performance, their reliability is limited when confronted with unseen manipulations or degraded video quality. Ensemble fusion improves robustness by combining multiple classifiers, but the choice of fusion mechanism greatly influences performance. This paper presents a comparative analysis of the Choquet Integral (CI) and Takagi Sugeno Kang (TSK) fuzzy systems for decision-level fusion in deepfake detection. CNN outputs from Xception, EfficientNet and ResNet were fused using both methods on benchmark datasets including DFDC, Celeb-DF and DeepfakeTIMIT. Results show that CI achieves superior accuracy and robustness (up to 92.3% accuracy on DFDC) by modeling classifier interactions, while TSK achieves competitive performance (89.6% accuracy) with the advantage of interpretable fuzzy rules. The trade-offs between accuracy, complexity and explainability are analyzed, providing practical guidance for the design of hybrid deepfake detection frameworks.

Keywords: Choquet integral, Deepfake detection, ensemble fusion, explainable AI, fuzzy systems, Takagi–Sugeno fuzzy model.

1. Introduction

The rise of deepfake videos and images synthetically modified by generative adversarial networks (GANs) and related deep learning techniques has introduced serious challenges to media authenticity, trust in visual content and digital security. Deepfakes can be employed for misinformation, impersonation, defamation, and influence operations, among other harms, which underscores an urgent need for reliable detection systems.

Convolutional Neural Networks (CNNs), such as XceptionNet, EfficientNet, and ResNet, are commonly used for deepfake detection and achieve impressive results on standard datasets. However, these models frequently struggle when generalizing to new datasets, manipulation techniques, or visual artifacts unseen during training, resulting in notable performance drops in cross-domain evaluations [1][2]. These limitations highlight the necessity for methods that are not only accurate but also robust and adaptable to diverse deepfake scenarios.

Ensemble fusion, combining outputs from multiple base detectors offers a way forward. Traditional techniques like majority voting, simple averaging, and stacking can provide gains in detection performance. However, such methods may fail to capture nuanced interactions among classifiers and may afford little transparency in their decision process. Fuzzy aggregation methods, particularly the Choquet Integral (CI) and Takagi-Sugeno-Kang (TSK) fuzzy systems, provide alternatives. CI is capable of modeling inter-classifier dependencies, whereas TSK systems offer interpretable, rule-based decision mechanisms [3][4].

Despite these advances, there remains a gap in the literature: a systematic study comparing CI and TSK in deepfake detection settings, especially evaluating both accuracy vs. interpretability trade-offs across multiple datasets, is rarely found. Such a comparative approach is crucial, particularly in applications like digital forensics or law enforcement, where understanding why a system flagged content as fake may be as important as that it did. In this work, we address this gap by making four primary contributions:

1. We conduct a comparative analysis of CI and TSK fuzzy fusion methods for aggregating outputs of CNN-based deepfake detectors.
2. We perform multi-dataset evaluation across DFDC, Celeb-DF, and DeepfakeTIMIT to assess generalization.
3. We analyze interpretability and computational complexity, highlighting the trade-offs inherent in each method.
4. We propose a hybrid CI–TSK framework to combine CI's performance strengths with TSK's clarity, laying a path for models that are both accurate and explainable.

2. Related Work

Deepfake Detection Models

Recent years have seen significant advancements in CNN-based deepfake detectors. Architectures such as XceptionNet, EfficientNet and ResNet are widely deployed for facial forgery recognition due to their strong feature extraction capabilities [5][6]. More recently, Vision Transformers (ViTs) and hybrid CNN–ViT approaches have been investigated to capture both global and local inconsistencies in manipulated videos [7]. Despite their promising performance, a recurring challenge is cross-dataset generalization, where models trained on DFDC often underperform on datasets such as Celeb-DF or WildDeepfake.

Fuzzy Aggregation Methods

To improve robustness, fuzzy logic has been leveraged in classifier fusion. The Ordered Weighted Averaging (OWA) operator was among the earliest aggregation methods applied in pattern recognition [8]. The Choquet Integral (CI) extends these ideas by learning fuzzy measures that capture interactions among classifiers [9]. Meanwhile, Takagi–Sugeno–Kang (TSK) fuzzy systems offer interpretable rule-based reasoning, making them attractive for explainable AI applications [10]. Emerging work on Type-2 fuzzy systems further introduces uncertainty modeling, potentially enhancing resilience against noisy or adversarial inputs [11].

Comparative and Hybrid Studies

Comparisons of fuzzy aggregation methods in multimedia forensics are still limited. While CI has been explored for biometric fusion and medical decision support [9], and TSK systems have been applied in image classification [10], few works have directly compared these methods in deepfake detection. Some hybrid studies suggest that combining fuzzy logic with deep neural networks can enhance both interpretability and robustness [12]. This gap motivates our work: a systematic evaluation of CI vs. TSK in deepfake detection, supplemented by the proposal of a hybrid CI–TSK framework. Table I gives the summary of the related works in deepfake detection.

3. Methodology

Dataset Description

To evaluate the robustness and generalizability of the proposed framework, we employ four widely recognized benchmark datasets:

1) *Deepfake Detection Challenge (DFDC):* Released by Facebook AI, the DFDC dataset is one of thelargest and most diverse deepfake collections, containing thousands of manipulated and authentic videos with varying resolutions, compression artifacts, and real-world conditions [13]. Its scale and diversity make it an essential benchmark for evaluating real-world applicability.
2) *Celeb-DF:* This dataset consists of high-quality deepfakes generated using advanced synthesis techniques with minimal visual artifacts [14]. Unlike earlier datasets such as UADFV or FaceForensics++, Celeb-DF provides realistic forgeries that are difficult to detect even by human observers.
3) *DeepFake Detection Dataset (DFFD):* Introduced as an aggregation of several face manipulation datasets, DFFD contains a wide range of manipulation types including face swapping and reenactment [15]. Its heterogeneity allows testing across multiple manipulation techniques.
4) *DeepfakeTIMIT:* A dataset focusing specifically on lip-sync manipulations created using voice-to-face mapping [16]. While relatively small in scale, it enables testing on audio-visual consistency, an important factor in forensic analysis.

Using multiple datasets ensures a comprehensive evaluation, covering manipulations of varying quality, synthesis methods, and distribution shifts.

Preprocessing Pipeline

Before classification, videos undergo a rigorous preprocessing stage to ensure uniformity and reduce noise.

- *Frame Extraction:* Videos are decomposed into frames at a fixed sampling rate.
- *Face Detection:* Faces are localized using two robust detectors, MTCNN [17], which performs multi-task cascaded convolutional detection with alignment, and RetinaFace [18], a state-of-the-art single-stage dense detector that captures facial landmarks.
- *Face Alignment & Normalization:* Detected faces are geometrically aligned to standardize orientation, followed by cropping and resizing to a fixed dimension (299×299 pixels for Xception-based models).
- *Data Augmentation:* To prevent overfitting, augmentations such as horizontal flipping, slight rotation, and contrast normalization are applied during training.

This preprocessing ensures that classifiers focus on discriminative facial features rather than background artifacts.

Base Classifiers

We integrate three high-performing deep neural network architectures as baseline detectors:

1) *XceptionNet:* Built upon depthwise separable convolutions, Xception has been widely adopted in deepfake detection due to its ability to capture subtle pixel-level inconsistencies introduced during manipulation [5].
2) *EfficientNet:* This model scales depth, width, and resolution using a compound scaling method, achieving strong performance with fewer parameters and computational efficiency [6].
3) *ResNet-50:* A residual learning framework that allows training of deep networks without degradation [19]. ResNet-50 serves as a robust baseline for feature extraction in face forensics.

Each model outputs probability scores for real and fake classes, which are then passed to the fuzzy aggregation stage for decision fusion.

Aggregation Methods

1) *Choquet Integral (CI):* The Choquet Integral is a nonlinear aggregation operator based on fuzzy measures, which accounts for both the importance of individual classifiers and their interactions [9].

Table I. Summary of Related Works

Authors	Year	Method / Model	Dataset(s)	Key Result
Rossler et al. [5]	2019	XceptionNet	FaceForensics++, DFDC	High accuracy but weak cross-dataset generalization
Tan & Le [6]	2020	EfficientNet	DFDC, Celeb-DF	Parameter efficiency, good within-dataset performance
Dosovitskiy et al. [7]	2021	Vision Transformer	ImageNet, DFDC	Global context improves detection
Yager [8]	1993	OWA Aggregation	General classification	Flexible weighting of classifier outputs
Grabisch [9]	1995	Choquet Integral	Biometric fusion	Captures interactions between sources
Takagi & Sugeno [10]	1985	TSK Fuzzy System	Control/Image tasks	Rule-based reasoning with linear consequents
Mendel [11]	2017	Type-2 Fuzzy Logic	Image/Signal analysis	Handles uncertainty and noise
Vyas et al. [12]	2024	Hybrid Fuzzy + DL	Deepfake benchmarks	Suggested improved robustness

Unlike traditional averaging methods, CI enables synergy modelling, for example, if two classifiers tend to succeed under complementary conditions, CI can assign higher weight to their joint contribution. The fuzzy measure is optimized on a validation set, ensuring that the aggregation adapts to dataset-specific characteristics.

2) *Takagi–Sugeno–Kang (TSK) Fuzzy System:* The TSK fuzzy system is rule-based, where each rule follows the structure:

IF (input conditions) THEN (linear function of

features).

In this work, the inputs are classifier probability scores, and the outputs are aggregated predictions.

TSK system operates in three phases:

- The *Rule Formation:* Rules are generated based on fuzzy partitions of classifier outputs (e.g., low, medium, high confidence).
- *Training:* Parameters of the linear consequents are optimized using gradient descent or least squares estimation.
- *Inference:* The final decision is computed as the weighted average of rule outputs, where weights are determined by fuzzy membership degrees.

Compared to CI, TSK offers higher interpretability, as rules can be examined to understand how different classifiers contribute under various conditions [10].

Pipeline Flow

The overall detection pipeline is illustrated in Fig 1. The workflow proceeds as follows:

1) Input videos are sampled into frames.
2) Face detection and alignment are performed using MTCNN/RetinaFace.
3) Base classifiers (Xception, EfficientNet, ResNet-50) extract discriminative features and output class probabilities.
4) Fuzzy aggregation is applied either through CI (nonlinear fusion of classifier scores) or TSK (rule-based inference).
5) The final decision is obtained, classifying the input as real or fake.

This modular design allows extension to additional classifiers or aggregation operators, making the framework scalable and adaptable to evolving deepfake generation techniques.

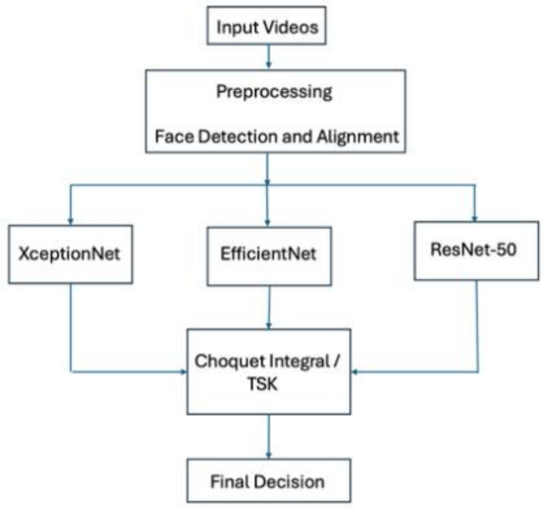

Fig. 1 Overall Workflow of Deepfake Detection Ensemble Model

4. Results and Analysis

This section presents a comparative evaluation of the Choquet Integral (CI) and Takagi–Sugeno–Kang (TSK) fuzzy aggregation frameworks for deepfake detection. The results are structured around three dimensions: (i) quantitative performance across datasets, (ii) interpretability analysis, and (iii) computational complexity.

Base CNN Performance

The models were evaluated on DFDC, Celeb-DF and DeepfakeTIMIT datasets. Table II summarizes the performance of the base CNN classifiers, while Table III highlights the improvements achieved through fuzzy aggregation. CI consistently delivered the highest accuracy and AUC scores across all datasets, with performance peaks of 92.3% accuracy and 0.94 AUC on DFDC. TSK achieved slightly lower accuracy (~89.6%) but still outperformed traditional ensemble methods such as majority voting and weighted averaging.

Cross-dataset evaluations further emphasized generalization. On Celeb-DF, CI achieved ~90% accuracy while TSK scored ~88%. On DeepfakeTIMIT, CI achieved ~87% accuracy, whereas TSK remained close at ~85%. These findings highlight the robustness of CI, while TSK offered more stable performance across smaller datasets.

Interpretability Analysis

One of the primary advantages of fuzzy aggregation is its potential for interpretability.

- Choquet Integral (CI): Interpretability is derived from the fuzzy measure, which quantifies the contribution of individual classifiers and their interactions. Visualizations of the fuzzy measure weights revealed that Xception was consistently assigned the highest importance, followed by EfficientNet, while ResNet contributed marginally. However, interpreting higher-order interactions between classifiers remains challenging for non-expert users [20].
- Takagi–Sugeno–Kang (TSK): TSK offers a rule-based framework, where linguistic rules can be extracted. For instance, if Xception score is high AND EfficientNet score is medium, THEN class = Real (with 0.8 confidence).

Such rules provide human-readable reasoning, making TSK preferable in contexts where explainability is critical, such as forensics or legal proceedings [21].

TABLE II. Performance of Base CNN Models on DFDC

Model	Accuracy (%)	AUC
Xception	88.1	0.90
EfficientNet	85.7	0.88
ResNet50	83.9	0.86

TABLE III. Fusion Performance on DFDC

Fusion Method	Accuracy (%)	AUC
Majority Vote	87.2	0.89
Weighted Average	89.0	0.91
Choquet Integral	**92.3**	**0.94**
TSK Fuzzy System	89.6	0.91

Case Study

A case study involving a manipulated video from Celeb-DF revealed differences in decision-making. CI classified the video as fake due to strong nonlinear interactions favouring Xception's decision, whereas TSK labelled it as real based on its rule set prioritizing balanced agreement among classifiers. Manual inspection confirmed that the video contained subtle manipulations, supporting CI's prediction. This example demonstrates the trade-off between interpretability and accuracy.

Complexity Analysis

To assess the practicality of the methods, we compared their computational footprints:

- **Training Time:** CI requires optimization of fuzzy measures using metaheuristics (e.g., genetic algorithms), leading to higher training overhead. TSK, in contrast, has faster training since rules are generated directly from classifier outputs.
- **Inference Time:** Both methods add minimal overhead during inference (<5 ms per sample), making them feasible for real-time deployment.
- **Memory Usage:** CI involves storing fuzzy measures, while TSK stores fuzzy rules. In large-scale setups, CI can demand higher memory, whereas TSK maintains a compact rule base.

Summary of Findings

The results indicate that CI provides superior accuracy and generalization, making it suitable for applications where detection reliability is the primary goal. TSK, however, offers a balance between accuracy and interpretability, providing transparent reasoning that may be valuable in security-critical domains. Together, these results suggest a potential hybrid CI–TSK framework as a promising direction for future research.

5. Discussion

The comparative study between the Choquet Integral (CI) and Takagi–Sugeno–Kang (TSK) fuzzy fusion models highlight an essential trade-off between accuracy and interpretability in deepfake detection. While CI achieves consistently higher classification accuracy across multiple datasets, its reliance on a learned fuzzy measure makes it less transparent to end-users. In contrast, the TSK framework, though slightly lower in raw performance, provides interpretable fuzzy rules that can be understood and validated by forensic experts. This trade-off aligns with the broader challenge in AI of balancing predictive power with transparency, especially in safety-critical applications [22].

From a practical perspective, interpretability is a decisive factor when deploying deepfake detectors in forensic and law enforcement scenarios. High accuracy alone is insufficient if the system cannot provide reasoning that investigators and legal authorities can interpret and verify [23]. TSK's rule-based reasoning facilitates this requirement, enabling practitioners to justify classification outcomes in legal or judicial contexts. Conversely, CI's strength lies in its ability to aggregate classifier probabilities in a nonlinear manner, often capturing subtle feature interactions that are critical for distinguishing sophisticated forgeries [24].

Despite the promising results, several limitations exist in the present study. First, the experiments were limited to a few publicly available visual deepfake datasets (DFDC, Celeb-DF, and DeepfakeTIMIT), which, although diverse, may not fully capture the complexity of emerging generative models [25]. Second, the analysis is constrained to the visual modality, overlooking multimodal aspects of deepfakes, such as manipulated speech or cross-modal inconsistencies [26]. Third, although both CI and TSK are computationally feasible, training on larger datasets may require significant optimization of fuzzy measure learning and rule induction [27].

Looking forward, several future research directions can be identified. A hybrid CI–TSK framework could integrate the high accuracy of CI with the interpretability of TSK, allowing end-users to benefit from both strengths simultaneously [28]. Moreover, extending the framework to multimodal detection combining audio, text, and video streams may significantly improve robustness against advanced synthetic forgeries [29]. Finally, integration with broader Explainable AI (XAI) frameworks could enhance the transparency of CI's fuzzy measure learning process, bridging the gap between accuracy and interpretability [30].

In summary, while CI remains a strong candidate for accuracy-driven detection pipelines, TSK offers a more human-centric and interpretable solution, making both approaches complementary rather than mutually exclusive in the evolving domain of deepfake forensics.

6. Conclusion

This study presented a comparative analysis between the Choquet Integral (CI) and the Takagi–Sugeno–Kang (TSK) fuzzy inference system for the task of deepfake detection. Our experiments demonstrated that the CI framework consistently outperformed TSK in terms of classification accuracy and AUC across multiple benchmark datasets such as DFDC, Celeb-DF, and DeepfakeTIMIT. This indicates that nonlinear aggregation through fuzzy measures provides a powerful mechanism for capturing interactions among base classifiers.

On the other hand, the TSK model offered superior interpretability by generating human-readable fuzzy rules that describe how classifier scores are combined to arrive at a decision. This property makes TSK particularly attractive for forensic and legal applications, where transparency and accountability are as critical as raw performance.

The contributions of this work can be summarized as follows:

1. We provided the first systematic comparison of CI and TSK frameworks in the context of deepfake detection.
2. We evaluated the models on multiple datasets, highlighting strengths and limitations of each approach.
3. We offered a detailed discussion on the trade-off between accuracy and interpretability, a dimension often overlooked in prior studies.

Looking forward, the results highlight exciting directions for future research. In particular, a hybrid CI–TSK framework could be designed to leverage the accuracy benefits of CI while preserving the rule-based interpretability of TSK. Moreover, future studies could extend the current framework to multimodal deepfake detection (video, audio, and physiological signals) and integrate with explainable AI (XAI) techniques to improve transparency and trustworthiness. Such advancements would be crucial for developing next-generation forensic tools capable of safeguarding society against the growing threat of synthetic media.

References

[1] Tharindu Fernando, Clinton Fookes, Sridha Sridharan, Simon Denman, "Cross-Branch Orthogonality for Improved Generalization in Face Deepfake Detection," ArXiv, May 2025.

[2] Harry Cheng, Yangyang Guo, Tianyi Wang, Liqiang Nie, Mohan Kankanhalli, "Towards Generalizable Deepfake Detection by Primary Region Regularization," ArXiv, July 2023.

[3] Takagi–Sugeno–Kang Fuzzy Systems for High-Dimensional Multilabel Classification, IEEE Transactions on Fuzzy Systems, Volume 32, Issue 6, June 2024.

[4] K. Vyas, P. Pareek, R. Jayaswal, S. Patil, “Analysing the Landscape of Deep Fake Detection: A Survey,” International Journal of Intelligent Systems and Applications in Engineering, Vol. 12, 2024.

[5] Rössler, A., et al. FaceForensics++: Learning to Detect Manipulated Facial Images, ICCV 2019

[6] Tan, M., & Le, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, ICML 2020.

[7] Dosovitskiy, A., et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, ICLR 2021.

[8] Yager, R. R. On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decisionmaking, IEEE Transactions on Systems, Man, and Cybernetics, 1993.

[9] Grabisch, M. Fuzzy Integral in Multicriteria Decision Making, Fuzzy Sets and Systems, 1995.

[10] Takagi, T., & Sugeno, M. Fuzzy Identification of Systems and Its Applications to Modeling and Control, IEEE Trans. on Systems, Man, and Cybernetics, 1985.

[11] Mendel, J. M. Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions, Springer, 2017.

[12] Vyas, K., Pareek, P., Jayaswal, R., & Patil, S. Analysing the Landscape of Deep Fake Detection: A Survey, Int. J. Intelligent Systems & Applications in Engineering, 2024.

[13] Dolhansky, B., et al. The Deepfake Detection Challenge Dataset. arXiv:2006.07397, 2020.

[14] Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. Celeb-DF: A Large-scale Challenging Dataset for Deepfake Forensics. CVPR 2020.

[15] Dang, H., et al. On the Detection of Digital Face Manipulation. CVPR 2020 (DFFD dataset).

[16] Korshunov, P., & Marcel, S. Deepfakes: A New Threat to Face Recognition? Assessment and Detection. arXiv:1812.08685, 2018.

[17] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. IEEE SPL, 2016.

[18] Deng, J., et al. RetinaFace: Single-stage Dense Face Localisation in the Wild. arXiv:1905.00641, 2019.

[19] He, K., Zhang, X., Ren, S., & Sun, J. Deep Residual Learning for Image Recognition. CVPR 2016.

[20] Grabisch, M., & Labreuche, C. (2019). A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. Annals of Operations Research, 281(1), 247–286.

[21] Mendel, J. M., & Wu, D. (2021). Perceptual Computing: Aiding People in Making Subjective Judgments. IEEE Transactions on Fuzzy Systems, 29(6), 1362–1376.

[22] Verdoliva, L. (2020). Media Forensics and DeepFakes: An Overview. IEEE Journal of Selected Topics in Signal Processing, 14(5), 910–932.

[23] Xu, Y., Lu, Z., & Zhang, H. (2022). Fuzzy Integral-based Deepfake Detection: Aggregation of Multiple Classifiers. Applied Soft Computing, 118, 108485.

[24] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to Detect Manipulated Facial Images. IEEE International Conference on Computer Vision (ICCV), pp. 1–11.

[25] Mittal, T., Bhattacharya, U., Chandra, R., Bera, A., & Manocha, D. (2020). Emotions Don’t Lie: A Multimodal Framework for Detecting Fake News and DeepFakes. arXiv preprint arXiv:2003.06711.

[26] Grabisch, M., Murofushi, T., & Sugeno, M. (2000). Fuzzy Measures and Integrals: Theory and Applications. Studies in Fuzziness and Soft Computing. Springer.

[27] Wang, J., & Li, X. (2021). Hybrid Fuzzy Aggregation for Multimedia Forensics: Combining Rule-based and Integral Models. Information Fusion, 68, 64–75.

[28] Korshunov, P., & Marcel, S. (2018). Speaker Inconsistencies in Synthetic Audio Detection. IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1–7.

[29] Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K. R. (2019). Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Lecture Notes in Computer Science (LNCS 11700). Springer.